



# Multivariate spatial meta kriging

Rajarshi Guhaniyogi<sup>a</sup>, Sudipto Banerjee<sup>b,\*</sup>

<sup>a</sup> Department of Applied Mathematics & Statistics, University of California Santa Cruz, SOE 2, 1156 High Street, Santa Cruz, CA 95064, USA

<sup>b</sup> Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095, USA



## ARTICLE INFO

### Article history:

Received 10 November 2017

Received in revised form 15 February 2018

Accepted 17 April 2018

Available online 4 May 2018

### Keywords:

Bayesian inference

Linear model coregionalization

Multivariate Gaussian process

Point-referenced data

Spatial meta kriging

Spatial stochastic process

## ABSTRACT

This work extends earlier work on spatial meta kriging for the analysis of large multivariate spatial datasets as commonly encountered in environmental and climate sciences. Spatial meta-kriging partitions the data into subsets, analyzes each subset using a Bayesian spatial process model and then obtains approximate posterior inference for the entire dataset by optimally combining the individual posterior distributions from each subset. Importantly, as is often desired in spatial analysis, spatial meta kriging offers posterior predictive inference at arbitrary locations for the outcome as well as the residual spatial surface after accounting for spatially oriented predictors. Our current work explores spatial meta kriging idea to enhance scalability of multivariate spatial Gaussian process model that uses linear model co-regionalization (LMC) to account for the correlation between multiple components. The approach is simple, intuitive and scales multivariate spatial process models to big data effortlessly. A simulation study reveals inferential and predictive accuracy offered by spatial meta kriging on multivariate observations.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

With the advent and expansion of Geographic Information Systems (GIS), along with related software, statisticians today routinely encounter large spatial datasets containing multiple variables observed across thousands of locations. Point-referenced data refer to situations where the variables of interest have been observed at fixed points or locations indexed by a coordinate system (e.g., Easting–Northing, Latitude–Longitude). Such data abound in the environmental and climate sciences where variables such as forest biomass, land-cover, temperature, precipitation and other such variables are studied. Such datasets are often studied via hierarchical models where the observed data are modeled as (partial) realizations of an underlying spatial stochastic process.

Hierarchical spatial Gaussian process models implemented through Markov chain Monte Carlo (MCMC) methods enable sophisticated analysis of such datasets. Likelihood computations for spatial process models typically entail matrix decompositions whose complexity increases in the order of  $O(n^3q^3)$  floating point operations (flops) in the number of observations,  $n$ , and the dimension of the outcome  $q$ . These operations are required in each iteration of the MCMC algorithm for parameter estimation making such multivariate spatial Gaussian process models to moderately large datasets (Cressie and Wikle, 2015; Banerjee et al., 2014).

There is a growing literature on the analysis of large spatial datasets mainly in the context of univariate spatial data. Briefly, these methods rely on “dimension-reduction” by approximating Gaussian process models with a low-rank model or with a sparse model. Low-rank processes are usually derived from expressing the Gaussian process using basis functions such as fixed-rank kriging (Cressie and Johannesson, 2008), or predictive processes and variants thereof (e.g., Banerjee et al., 2008;

\* Corresponding author.

E-mail address: [sudipto@ucla.edu](mailto:sudipto@ucla.edu) (S. Banerjee).

Finley et al., 2009; Guhaniyogi et al., 2011). Sparse models posit that spatial correlation between two distantly located observations is nearly zero, so little information is lost by assuming conditional independence given the intermediate locations. A few important classes of sparse models have emerged recently, such as models based covariance tapering (e.g., Furrer et al., 2006; Kaufman et al., 2008; Du et al., 2009; Shaby and Ruppert, 2012), composite likelihoods (e.g., Eidsvik et al., 2014) or nearest neighbor models (e.g., Vecchia, 1988; Rue et al., 2009; Stein et al., 2004; Datta et al., 2016). There are approaches that extend low-rank or sparse processes to enhance scalability of multivariate spatial data (Banerjee et al., 2010; Guhaniyogi et al., 2013; Guhaniyogi, 2017), though full Bayesian computations with low-rank multivariate processes become prohibitive with more than 5000–10,000 observations. In general, considerably less attention has been given in the literature to boost scalability of multivariate spatial processes which identify complex association not only across the spatial domain but also across its components over space.

We exploit parallelization by dividing the entire dataset into a moderately large number of subsets, where each subset consists of a tractable number of locations, and fit multivariate spatial regression models independently to each of these subsets. Thus, in a multicore computing architecture, each subset is analyzed on a CPU dedicated to itself. We only retain the converged posterior samples from each subset on its respective CPU. The posteriors from various subsets (also known as “subset posteriors”) are then combined optimally to yield a single posterior distribution (the “meta-posterior”) for the model parameters that substitutes the full posterior distribution for all inferential and predictive tasks. To combine subset posteriors, we adapt the notation of a geometric median of subset posteriors (see, e.g., Minsker et al., 2014; Guhaniyogi and Banerjee, in press). Though the idea has been proposed in the context of predictive models for independent data (Minsker et al., 2014), or for scaling univariate spatial process models (Guhaniyogi and Banerjee, in press), this article extends the technique to multivariate spatial models. The approach simply demands obtaining posterior samples for the process parameters and spatial random effects from a multivariate spatial model fitted on multiple subsets in parallel followed by deriving the meta-posterior. We coin this method as the “multivariate spatial meta kriging (MSMK)”.

The remainder of the article evolves as follows. Section 2 discusses the concept of multivariate spatial meta kriging, while Section 3 demonstrates its empirical performance with a brief simulation study. Finally, Section 4 concludes the short paper with some discussion and general conclusions.

## 2. Pooled Bayesian inference for multivariate spatial process models

Let  $\mathcal{D} \subset \mathbb{R}^2$  denote our spatial domain of interest and let  $s$  be a generic point in  $\mathcal{D}$ . Geostatistical multivariate settings customarily assume, for each location  $s \in \mathcal{D}$ , a linear model

$$y(s) = X(s)' \beta + w(s) + \epsilon(s), \quad (1)$$

where  $y(s) = (y_1(s), y_2(s), \dots, y_q(s))'$  is a  $q \times 1$  response vector,  $X(s)'$  is a  $q \times p$  matrix of spatially referenced predictors (or completely known functions), whose  $i$ th row is a  $1 \times p$  vector of spatially referenced predictors  $x_i(s)'$ , capturing large-scale variation or trends and  $w(s) = (w_1(s), w_2(s), \dots, w_q(s))'$  is a  $q \times 1$  (multivariate) spatial process, modeling local spatial patterns and structured dependence.

The customary process specification for  $w(s)$  is a zero-centered  $q$ -variate Gaussian process. The process  $w(s)$  is completely specified by its cross-covariance function  $C_w(s, t; \alpha)$ , which, for any pair of locations  $s$  and  $t$ , is a  $q \times q$  matrix with  $\text{cov}\{w_{l_1}(s), w_{l_2}(t)\}$  as its  $(l_1, l_2)$ th element and  $\alpha$  is a collection of process parameters. Therefore,  $C_w(s, s; \alpha)$  is precisely the variance-covariance matrix for the elements of  $w(s)$  within site  $s$ . For a detailed review of how to construct valid cross-covariance matrix function, see Gelfand et al. (2010), Gneiting et al. (2010) and Guhaniyogi et al. (2013). This article considers one popular way to construct the cross-covariance matrix by a linear model co-regionalization (LMC).

Assume,  $v_l(s)$ 's are independent univariate Gaussian processes with mean 0 and covariance function  $\xi(s, t; \theta_l)$ . The LMC approach (Gelfand et al., 2004) proposes that  $w(s) = Av(s)$ ,  $v(s) = (v_1(s), \dots, v_q(s))'$ , which yields a highly structured cross-covariance function,

$$C_w(s, t; \alpha) = AC_v(s, t; \alpha)A' = \sum_{l=1}^q a_l a_l' \xi(s_1, s_2; \theta_l), \quad (2)$$

where  $a_l$  is the  $l$ th column of  $A$ ,  $\alpha = \{\theta_1, \dots, \theta_q\}$  (in the univariate case  $\alpha = \theta_1$ ). One natural choice of  $\xi(\cdot, \cdot)$  comes from the class of Matérn correlation functions,

$$\xi(s, t; \theta) = \frac{1}{2^{\theta_2-1} \Gamma(\theta_2)} (\|s - t\| \theta_1)^{\theta_2} \mathcal{K}_{\theta_2}(\|s - t\|; \theta_1); \quad \theta_1 > 0, \theta_2 > 0, \quad (3)$$

where  $\theta = \{\theta_1, \theta_2\}$ ,  $\theta_1$  and  $\theta_2$  are decay and smoothness parameters, respectively. This implies that for any finite set of  $n$  locations,  $\{s_1, s_2, \dots, s_n\}$ , the  $nq \times 1$  vector of realizations,  $w = (w(s_1)', w(s_2)', \dots, w(s_n'))'$  follows a multivariate normal distribution with zero mean and a  $nq \times nq$  blocked covariance matrix  $C_w(\alpha)$  whose  $(i, j)$ th block is given by the  $q \times q$  matrix  $C_w(s_i, s_j; \alpha)$ .  $\epsilon(s) = (\epsilon_1(s), \dots, \epsilon_q(s))'$  is the multivariate unstructured error with  $\epsilon_l(s) \stackrel{\text{ind}}{\sim} N(0, \tau_l^2)$ .

Given a set of locations  $S = \{s_i : i = 1, 2, \dots, n\}$ , where  $y(s)$  and  $X(s)$  have been observed, the spatial regression in (1) is extended to a hierarchical linear mixed model framework

$$y = X\beta + w + \epsilon, \quad \epsilon \sim N(0, I \otimes \Psi), \quad (4)$$

where  $y = (y(s_1)', \dots, y(s_n'))'$ ,  $w = (w(s_1)', \dots, w(s_n'))'$ ,  $w \sim N(0, C_w(\alpha))$ , and  $\epsilon = (\epsilon(s_1)', \dots, \epsilon(s_n'))'$  are  $nq \times 1$  vectors,  $X$  is the  $nq \times p$  matrix of regressors ( $p < n$ ) with  $X(s_i)$  as its  $i$ th block row of dimension  $q \times p$ ,  $\Psi$  is a  $q \times q$  diagonal matrix given by  $\text{diag}(\tau_1^2, \dots, \tau_q^2)$ ,  $\beta \sim N(\mu_\beta, \Sigma_\beta)$  is the prior distribution for the slope vector,  $\mu_\beta$  and  $\Sigma_\beta$  are assumed fixed, and  $\alpha$  is assigned a proper prior distribution  $p(\alpha)$ .

Often, the parameter space is shrunk by integrating out  $w$  from (4). Let  $\Omega$  denote the parameter space for our model of interest. Thus,  $\Omega = \{\beta, w, \alpha, \Psi\}$  for (4) or  $\Omega = \{\beta, \alpha, \Psi\}$  for the model with  $w$  integrated out of (4). Bayesian inference proceeds, customarily, by sampling  $\Omega$  from (4), or its marginalized counterpart, using Markov chain Monte Carlo (MCMC) methods. Irrespective of the specific parametrization or estimation algorithm, model fitting usually involves matrix decompositions for  $C_w(\alpha)$  requiring  $\sim (nq)^3$  floating point operations (flops) and  $\sim (nq)^2$  memory units in storage. These become prohibitive for large  $n$  since  $C_w(\alpha)$ , in general, has no exploitable structure.

Let  $S$  be partitioned into  $K$  exhaustive and mutually exclusive subsets  $S_1, \dots, S_K$ . Let  $\{y_k, X_k\}$  be the corresponding data partitions, for  $k = 1, 2, \dots, K$ , where  $y_k$  is  $n_k q \times 1$  and  $X_k$  is  $n_k q \times p$ . Assume that we have collected posterior samples for  $\Omega$  from (4) applied independently to each of  $K$  subsets of the data. Let  $\Pi(\Omega | y_k)$  be the posterior distribution of  $\Omega$  from the  $k$ th subset, referred to as the  $k$ th subset posterior. Assume that  $\Omega_{k,1}, \dots, \Omega_{k,P}$  are the  $P$  post burn-in posterior samples from  $\Pi(\Omega | y_k)$ . Following Guhaniyogi and Banerjee (in press), we propose to combine the  $\Pi(\Omega | y_k)$ 's to arrive at a legitimate probability density  $\Pi_M(\Omega | y)$ , referred to as the “meta-posterior”.

The approach computes the unique Geometric Median (GM) (Minsker et al., 2014)  $\Pi_M$  of  $\Pi(\cdot | y_k)$ 's. Let  $\Pi_k \equiv \Pi(\Omega | y_k)$  resides on a Banach space  $\mathcal{H}$  equipped with norm  $\|\cdot\|$ . The GM is defined as

$$\pi_M(\cdot | y) = \arg \min_{\Pi \in \mathcal{H}} \sum_{k=1}^K \|\Pi_k - \Pi\|_\rho. \quad (5)$$

The norm quantifies the distance between any two posterior densities  $\Pi_1(\cdot)$  and  $\Pi_2(\cdot)$  as

$$\|\Pi_1 - \Pi_2\|_\rho = \left\| \int \rho(\Omega, \cdot) d(\Pi_1 - \Pi_2)(\Omega) \right\|,$$

where  $\rho(\cdot)$  is a positive-definite kernel function. In what follows, we assume  $\rho(z_1, z_2) = \exp(-\|z_1 - z_2\|^2)$ . We follow the same algorithm outlined in Guhaniyogi and Banerjee (in press) to compute  $\Pi_M$ . In particular, the algorithm computes weights  $\zeta_k(y) \in (0, 1)$  corresponding to the  $k$ th subset posterior  $\Pi_k$  so that  $\Pi_M = \sum_{k=1}^K \zeta_k(y) \Pi_k$ . The approach has similarities with Bayesian meta analysis and is coined as the “multivariate spatial meta kriging (MSMK)”.

Guhaniyogi and Banerjee (in press) argues  $\Pi_M$  to be a theoretically justifiable approximation to the full posterior. Therefore, for the purpose of prediction and surface interpolation, the full posterior can be replaced with  $\Pi_M$ . To be more precise, let the posterior predictive distribution at a location  $s_0$  be given by  $\Pi(y(s_0) | y)$  while the subset posterior predictive distributions be denoted by  $\Pi(y(s_0) | y_k)$ . Following the discussion above, an approximation to  $\Pi(y(s_0) | y)$  is  $\sum_{k=1}^K \zeta_k(y) \Pi(y(s_0) | y_k)$ . An empirical approximation to the posterior predictive distribution is thus given by  $\sum_{k=1}^K \sum_{m=1}^P \frac{\zeta_k(y)}{P} 1_{y(s_0)^{(m,k)}}$ , where  $y(s_0)^{(m,k)}$  is the  $m$ th posterior predictive sample from the  $k$ th subset. Posterior predictive summaries such as the median and a 95% predictive interval for the unobserved  $y(s_0)$  are readily available from this empirical distribution. Posterior predictive mean and 95% credible interval of the spatial surface  $w(s_0)$  at an unobserved location  $s_0$  are analogously obtained.

### 3. Empirical study

This section presents an empirical performance of MSMK in terms of parameter estimation, surface interpolation and prediction at unobserved locations. In the simulation example, we randomly divide the data into  $K$  exhaustive and mutually exclusive subsets. The *random partitioning* (RP) scheme creates subsets with points from every subregions of the domain. Alternatively, one can cluster full data to  $K$  different clusters using the  $k$ -means clustering and use each of these clusters as a subset. Later we demonstrate that this  $k$ -means clustering (KM) of locations into subsets leads to inferior inference than MSMK fitted on subsets constructed with the random partitioning scheme. A more sophisticated approach would be to partition the domain into sub-domains and include representative samples from each sub-domain in a subset. We refer to this partitioning scheme as *random-block partitioning* (RBP) and show its indistinguishable performance with random partitioning. All these evidences with brief discussions are provided at the end of Section 3.1. MSMK is run in R with multicore parallelization implemented through `foreach` and `doParallel` packages available in the Comprehensive R Archive Network (CRAN: <https://cran.r-project.org/>).

#### 3.1. Analysis of synthetic data

To illustrate the performance of the MSMK, 20,500 bivariate observations within a unit square domain are generated from the likelihood given by (4), out of which 20,000 are used for model fitting and the rest 500 for prediction. The matrix of predictors  $X(s)$  at any location is taken to be a  $2 \times 4$  matrix with first row  $x_1(s) = (1, z_1(s), 0, 0)'$ , second row  $x_2(s) = (0, 0, 1, z_2(s))'$  where  $z_1(s), z_2(s)$  are drawn independently from  $N(0, 1)$ , and the corresponding coefficient vector is  $\beta = (\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12})' = (1, 2, 0.4, 1)'$ . An exponential spatial correlation function was assumed for all spatial processes,

**Table 1**

The median and 95% Bayesian credible intervals of parameters for MSMK for  $K = 20, 40$ . As a competitor, SMK is also fitted independently to the two components for  $K = 20, 40$ . Also presented are mean squared prediction error (MSPE), interval score (INT) and prediction interval coverage (CVG) for the two competitors over 500 out of sample observations. Computation time in minutes is also presented.

	True	MSMK( $K = 20$ )	SMK( $K = 20$ )	MSMK( $K = 40$ )	SMK( $K = 40$ )
$\beta_{01}$	1	2.04 (−0.54, 4.12)	1.95 (0.73, 3.06)	1.93 (0.25, 3.66)	1.86 (0.02, 3.65)
$\beta_{11}$	2	1.97 (1.96, 2.02)	1.98 (1.95, 2.03)	1.99 (1.95, 2.04)	2.00 (1.93, 2.06)
$\beta_{02}$	0.4	0.97 (−0.37, 2.33)	0.98 (−0.19, 1.94)	1.03 (−0.17, 2.27)	1.07 (−0.15, 2.13)
$\beta_{12}$	1	1.00 (0.97, 1.02)	0.99 (0.98, 1.02)	0.99 (0.94, 1.05)	0.99 (0.95, 1.04)
$\tau_1^2$	0.2	0.03 (0.02, 0.04)	0.01 (0.007, 0.02)	0.04 (0.05, 0.08)	0.02 (0.01, 0.04)
$\tau_2^2$	0.1	0.02 (0.02, 0.03)	0.01 (0.008, 0.01)	0.04 (0.03, 0.05)	0.01 (0.01, 0.03)
$A_{11}$	2	1.80 (1.16, 2.40)	–	1.55 (1.14, 2.2)	–
$A_{12}$	0.9	0.94 (0.78, 1.23)	–	0.82 (0.60, 1.19)	–
$A_{22}$	1	0.89 (0.54, 1.15)	–	0.89 (0.76, 0.95)	–
$\phi_1$	2	1.96 (1.20, 4.69)	2.35 (1.12, 3.51)	2.46 (1.23, 4.47)	2.64 (1.29, 4.84)
$\phi_2$	4	3.89 (1.76, 5.77)	3.74 (2.12, 5.35)	3.55 (2.07, 5.61)	3.36 (1.89, 5.29)
MSPE	–	0.18	0.22	0.26	0.27
CVG	–	0.99	0.98	0.99	0.98
INT	–	1.65	1.76	1.99	2.08
Time(in min)	–	116.46	26.82	29.07	5.98

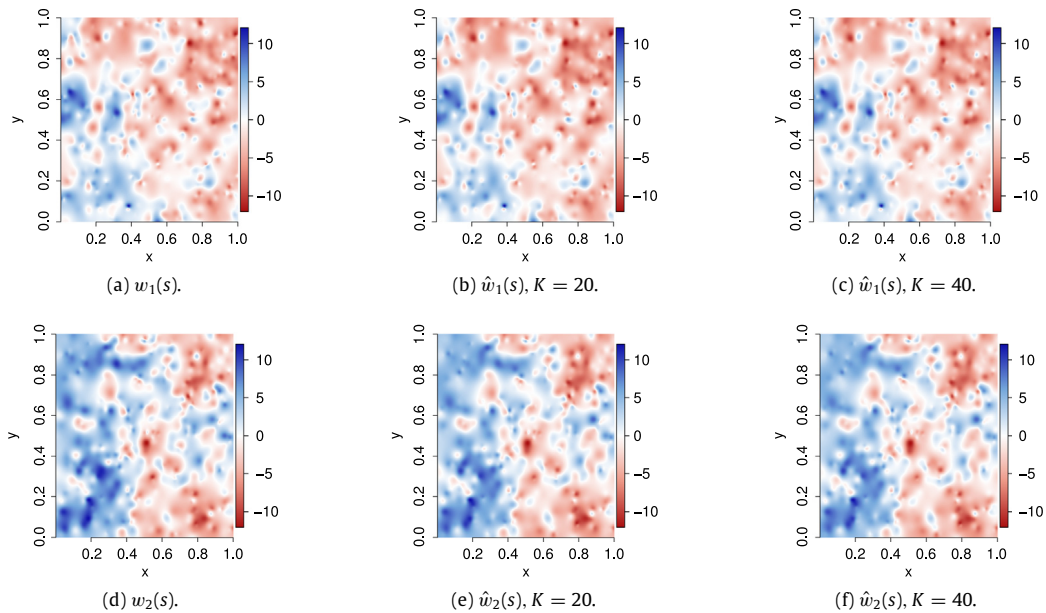
i.e.,  $\theta_2$  was fixed at 0.5 in (3). For the sake of identifiability of each element of  $A$ , we assume  $A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ , with  $A_{11} > 0$  and  $A_{22} > 0$ . The matrix  $A$ , then, identifies with the lower-triangular Cholesky factor of the covariance matrix for  $w(s)$ . It is assumed to be stationary across locations. The column labeled *True* in Table 1 contains the parameter values used to generate the data.

As a competitor to MSMK, we also fit SMK independently on the two components for the simulated data to gauge the inferential benefits fetched by MSMK. For estimating the LMC model in each subset for both competitors, we assign a flat prior to the common intercept  $\beta$  and  $N(0, 1)$  prior for  $A_{21}$ . All variance parameters  $\{\tau_1^2, \tau_2^2, A_{11}, A_{22}\}$  were assigned  $IG(2, 1)$  (mean = 1) prior. Decay parameters  $\phi_1$  and  $\phi_2$  in the exponential correlation functions for the two components of  $w(s)$  are assigned  $U(1, 8)$  prior that gives fairly wide support for range parameters given that the maximum inter-location distance in the generated data is 1.36. Both competitors are fitted with  $K = 20, 40$  subsets.

Table 1 presents posterior means along with 95% credible intervals for all the parameters. All parameters are estimated accurately by both MSMK and SMK except the nugget parameters,  $\tau_1^2$  and  $\tau_2^2$ , which are found to be somewhat overestimated by MKSMK. However, we do not believe this is a consequence of the MKSMK approach. Instead, this is likely an artifact of the multivariate model itself, where estimation of the (weakly identifiable) nugget gets affected by highly correlation between the two component processes (reflected by  $A_{21} = 0.9$ ). Turning to robustness, the parameter estimates from MSMK are very robust when the number of subsets varies within a certain range. However, with increasing number of subsets, the predictive performance of MSMK deteriorates, as seen by comparing the mean squared prediction error (MSPE) values of MSMK between  $K = 20$  and  $K = 40$ . This observation is further reinforced by the coverage of the 95% predictive interval (CVG) and the interval score (INT) measure (Gneiting and Raftery, 2007) of the 95% predictive intervals. With increasing number of subsets, the 95% predictive interval shows an upward trend in terms of the INT score keeping the same CVG measure. Turning to model comparisons, joint modeling of bivariate components using MSMK yields significant advantages in terms of prediction for  $K = 20$ . However, such an advantage becomes more bleak as the number of subsets increases. With smaller subsets the correlation between the two components has little effect on prediction, thus causing MSMK to behave similarly to SMK.

A pictorial depiction of the residual spatial surfaces for outcomes 1 and 2 from MSMK with  $K = 20, 40$  can be found in Fig. 1. Comparing these estimated residual surfaces with the true data generating surfaces, it is evident that the local spatial variability is adequately captured by spatial surfaces in MSMK. We have compared our results also with the residual surfaces constructed from the modified predictive process, though the results are not shown due to the space constraint. Needless to say, the residual surfaces constructed from the modified predictive process model are grossly oversmoothed. Finally, we note that the computational complexity of MSMK is dominated by  $(mq)^3$  floating point operations which leads to manageable run times even with a non-optimized implementation in R.

Finally, Table 2 presents MSPE, CVG and INT of 95% predictive intervals for MSMK corresponding to random partitioning, random block partitioning and k-means clustering of points in the construction of subsets, as described before. The table suggests effectively indistinguishable performance for random partitioning and random block partitioning, while k-means clustering shows inferior performance. Intuitively, both random partitioning and random block partitioning lead to subset posteriors which are noisy approximations to the full posterior. On the contrary, k-means clustering of points constructs subsets which are not represented by points from the entire domain. Therefore all empirical performances are shown only for the random partitioning scheme.



**Fig. 1.** Row 1 gives true and estimated spatial surfaces (posterior mean) for the first outcome of MSMK for  $K = 20, 40$ . Row 2 provides the same for the second component of MSMK.

**Table 2**

Mean squared prediction error (MSPE), interval score (INT) and prediction interval coverage (CVG) of 95% predictive interval for the three data subsetting schemes over 500 out of sample observations.

	$K = 20$			$K = 40$		
	MSPE	CVG	INT	MSPE	CVG	INT
RP	0.18	0.99	1.65	0.26	0.99	1.99
KM	2.21	0.99	5.77	3.01	0.99	5.76
RBP	0.17	0.99	1.65	0.26	0.99	1.96

#### 4. Conclusion & future work

This article extends spatial meta kriging to multivariate spatial data. We propose partitioning the data, fitting multivariate Gaussian process models to each subset, and then combining inference across subsets. The proposed idea offers a theoretically justifiable framework to scale multivariate spatial process models to massive numbers of observations in manageable computation time. Imminent future work will focus on employing spatial meta-kriging or its variants to seamlessly scale low-rank or sparse models to millions of observations. There is also a scope of utilizing the power of the spatial meta-kriging approach to enhance scalability for various spatio-temporal models.

A potential concern with the current specification of MSMK is that if the underlying spatial process has substantial non-stationary local behavior, then the subset posteriors are likely to miss important local behavior if samples are sparsely drawn from each subset and this will lead to the MSMK missing some local non-stationary features as well. While one can increase the number of data points in each subset to improve MSMK's performance, this will detract from the computational gains and perhaps preclude applying MSMK to infer from massive datasets exhibiting non-stationary behavior. We have reported a similar issue in the univariate SMK paper [Guhaniyogi and Banerjee \(in press\)](#) and recently resolved this issue with a slightly different divide and conquer approach for univariate non-stationary Gaussian process models ([Guhaniyogi et al., 2017](#)). Extension to such an approach for the multivariate non-stationary models requires thorough investigation which we plan to commit in the near future. We also plan to theoretically explore the upper bound on the number of subsets as a function of the sample size to obtain desirable inference.

Finally, one wonders whether multivariate observations can be split into different subsets. Note that the theory for MSMK only supports splitting along the samples. Splitting multivariate observations will require a whole new theory and algorithms for combining subset posteriors; we intend to pursue this in a different article. Additionally, due to the space constraints, we have not been able to pursue a more elaborate simulation study with multivariate observations having a few coordinates missing for a few sample points. This, too, should be addressed in future work.

## Acknowledgments

The second author was supported by grants National Science Foundation grants DMS-1513654 and IIS-1562303 and by NIH grant NIEHS 1R01ES027027-01.

## References

- Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. *Hierarchical Modeling and Analysis for Spatial Data*. Crc Press.
- Banerjee, S., Finley, A.O., Waldmann, P., Ericsson, T., 2010. Hierarchical spatial process models for multiple traits in large genetic trials. *J. Amer. Statist. Assoc.* 105 (490), 506–521.
- Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 70 (4), 825–848.
- Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 70 (1), 209–226.
- Cressie, N., Wikle, C.K., 2015. *Statistics for Spatio-Temporal Data*. John Wiley & Sons.
- Datta, A., Banerjee, S., Finley, A.O., Gelfand, A.E., 2016. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* 111 (514), 800–812.
- Du, J., Zhang, H., Mandrekar, V., 2009. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* 37 (6A), 3330–3361.
- Eidsvik, J., Shaby, B.A., Reich, B.J., Wheeler, M., Niemi, J., 2014. Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.* 23 (2), 295–315.
- Finley, A.O., Banerjee, S., Waldmann, P., Ericsson, T., 2009. Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial datasets. *Biometrics* 65 (2), 441–451.
- Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* 15 (3), 502–523.
- Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M. (Eds.), 2010. *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL.
- Gelfand, A.E., Schmidt, A.M., Banerjee, S., Sirmans, C., 2004. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* 13 (2), 263–312.
- Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* 105 (491), 1167–1177.
- Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* 102 (477), 359–378.
- Guhaniyogi, R., 2017. Multivariate bias adjusted tapered predictive process models. *Spatial Stat.* 21, 42–65.
- Guhaniyogi, R. and Banerjee, S., (in press). Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics*. <https://doi.org/10.1080/00401706.2018.1437474>.
- Guhaniyogi, R., Finley, A.O., Banerjee, S., Gelfand, A.E., 2011. Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics* 22 (8), 997–1007.
- Guhaniyogi, R., Finley, A.O., Banerjee, S., Kobe, R.K., 2013. Modeling complex spatial dependencies: Low-rank spatially varying cross-covariances with application to soil nutrient data. *J. Agric. Biol. Environ. Stat.* 18 (3), 274–298.
- Guhaniyogi, R., Li, C., Savitsky, T.D., Srivastava, S., 2017. A divide-and-conquer Bayesian approach to large-scale kriging, arXiv preprint. arXiv:1712.09767.
- Kaufman, C.G., Schervish, M.J., Nychka, D.W., 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* 103 (484), 1545–1555.
- Minsker, S., Srivastava, S., Lin, L., Dunson, D., 2014. Scalable and robust Bayesian inference via the median posterior. In: *Proceedings of the 31st International Conference on Machine Learning, ICML-14*, pp. 1656–1664.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71 (2), 319–392.
- Shaby, B., Ruppert, D., 2012. Tapered covariance: Bayesian estimation and asymptotics. *J. Comput. Graph. Statist.* 21 (2), 433–452.
- Stein, M.L., Chi, Z., Welty, L.J., 2004. Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (2), 275–296.
- Vecchia, A.V., 1988. Estimation and model identification for continuous spatial processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 50 (2), 297–312.